

Глава 18. Множественные таблицы данных Power Pivot с различной степенью детализации

Это продолжение перевода книги Роб Колли. Формулы DAX для Power Pivot. Главы не являются независимыми, поэтому рекомендую начать сначала.

[Предыдущая глава](#) [Содержание](#) [Следующая глава](#)

Ваши таблицы данных могут иметь различную структуру:

- В детализированных наборах данных каждая строка содержит относительно немного информации. Высокодисперсные наборы данных, как правило, содержат много коротких строк.
- В менее детализированных наборах данных каждая строка представляет собой «более толстый» фрагмент данных, и поэтому эти наборы данных, как правило, содержат меньше строк.
- Если ваши таблицы данных имеют одинаковую степень детализации, предыдущая глава – это все, что вам нужно. Но если у них разные степени детализации, вам пригодятся идеи, содержащиеся в этой главе.
- В общем случае, чем более дисперсен набор данных, тем более мощной и гибкой может быть ваша модель.

	CalendarYear	MonthNumberOfYear	SalesTerritoryReg...	Subcateg...	Budgeted Sales
1	2001		7 Australia	Mountain Bikes	\$71 510
2	2001		7 Australia	Road Bikes	\$190 248
3	2001		7 Canada	Mountain Bikes	\$4 183
4	2001		7 Canada	Road Bikes	\$15 429
5	2001		7 France	Mountain Bikes	\$7 916
6	2001		7 France	Road Bikes	\$31 825
7	2001		7 Germany	Mountain Bikes	\$4 384
8	2001		7 Germany	Road Bikes	\$36 068
9	2001		7 Northwest	Mountain Bikes	\$12 058
10	2001		7 Northwest	Road Bikes	\$49 089
11	2001		7 Southwest	Mountain Bikes	\$3 059
12	2001		7 Southwest	Road Bikes	\$59 190
13	2001		7 United Kingdom	Mountain Bikes	\$5 893
14	2001		7 United Kingdom	Road Bikes	\$28 572
15	2001		8 Australia	Mountain Bikes	\$51 200

Рис. 18.1. Бюджет продаж, импортированный в модель данных Power Pivot

Бюджет и факт

У вас есть таблица продаж (фактические данные), где каждая строка представляет отдельную транзакцию. В нашем примере это 60 тысяч строк. У вас также есть бюджет (план продаж), менее детализированный (рис. 18.1). Как правило, нас интересует, как наши продукты продаются по сравнению с бюджетом?

Решение этой проблемы в обычном Excel утомительно. Решение включает в себя создание двух сводных – одной для продаж, другой для бюджета, а затем написание формул, которые индексируют каждую сводную, чтобы сформировать единый отчет «продажи по сравнению с бюджетом». Требуется некоторое время, чтобы подготовить такой отчет. А затем, когда кто-то с неизбежностью захочет увидеть чуть-чуть другой формат или уровень детализации, вы потратите почти столько же времени, как и в первый раз...

В Power Pivot продажи и бюджет могут сосуществовать в одной и той же сводной таблице. При том что в модели данных это будут две разные таблицы. Всё что нам нужно, это создать «правильные» связи. Если это было совсем просто, то не понадобилась бы отдельная глава. К сожалению, нас подстерегает проблема: таблица [Budget] не связывается ни с одной из наших таблиц поиска. Например, попытка связать [Budget] с [Products], используя единственный общий столбец – [Subcategory] (рис. 18.2), приводит к ошибке «многие ко многим».

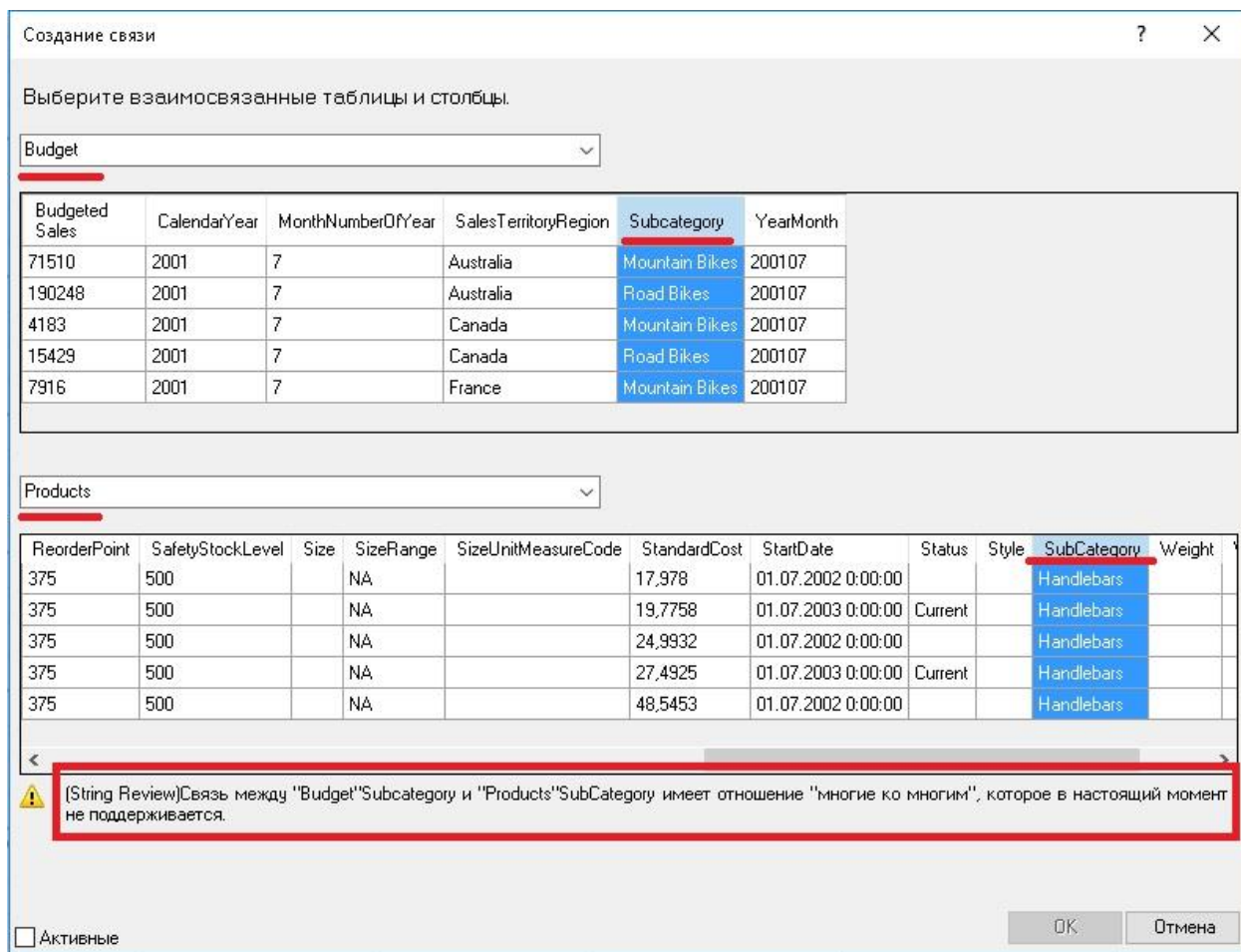


Рис. 18.2. Попытка связать бюджет с продуктами... приводит к ошибке «многие ко многим»

И с этим ничего нельзя поделать, так как каждое значение подкатегории (например, «горные велосипеды») появляется много раз в каждой таблице.

Нам нужна таблица поиска подкатегорий – [Subcategory], которая будет содержать только уникальные записи (рис. 18.3).

	SubCat
1	Bib-Shorts
2	Bike Racks
3	Bike Stands
4	Bottles and Cages
5	Bottom Brackets
6	Brakes
7	Caps
8	Chains
9	Cleaners
10	Cranksets
11	Derailleurs

Рис. 18.3. Новая таблица поиска [Subcategory]

Затем мы связываем таблицу [Subcategory] с таблицами [Products] и [Budget]. Теперь наша диаграмма выглядит так:

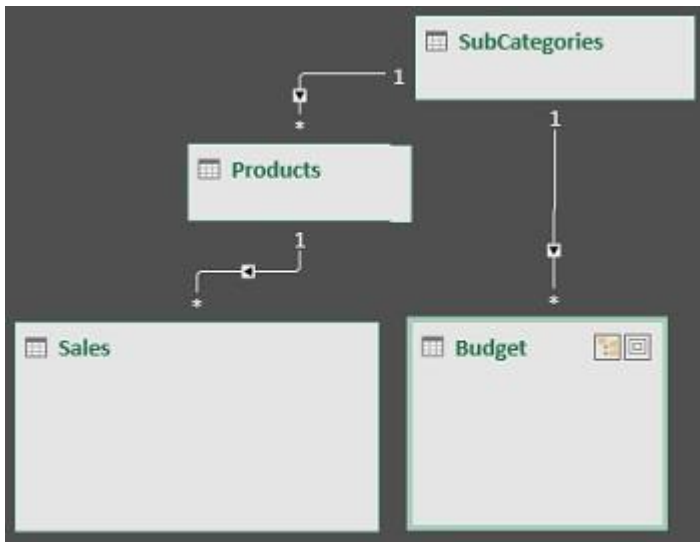


Рис. 18.4. Фрагмент модернизированной модель данных

Контекст фильтра перетекает из подкатегорий в продукты, а затем из продуктов в продажи.

Проще всего создать новую таблицу поиска [Subcategory] с помощью Power Query. Вы можете автоматически создать уникальный список значений подкатегорий, используя запрос к базе данных и преобразовав данные в [Power Query](#) (чью команду *Удалить дубликаты* также можно назвать *Создать таблицу поиска из Таблицы данных*). Это избавит вас от ручного обновления в будущем, если появятся новые подкатегории (или удаляться старые).

Поскольку детализация бюджета по датам сводится только к парам год/месяц, нам нужна новая таблица поиска с той же степенью детализации (рис. 18.5).

[YearMonth]	fx	
Year	Month	YearMonth
2001	10	200110
2001	11	200111
2001	12	200112
2001	7	200107
2001	8	200108
2001	9	200109

Years | SubCategories | CompetitorSales | CountTest | **YearMonths** | Cu

Запись: 1 из 37

Рис. 18.5. Новая таблица поиска [YearMonths]

[YearMonths] – вычисляемый столбец. Функция `FORMAT()` используется для добавления нуля перед однозначными числами месяцев. Это не обязательно, но для правильной сортировки значений в столбце [YearMonths] очень полезно.

Мы добавляем такой же вычисляемый столбец в таблицы [Budget] и [Calendar], а затем создаем связи таблицы [YearMonths] с таблицами [Budget] и [Calendar]. Для таблицы [SalesTerritory] нам не нужно создавать новую таблицу поиска. Здесь [Budget] соответствует уже имеющейся детализации таблицы [SalesTerritory]. Поэтому мы просто создаем связь между этими таблицами. Так как таблиц много, отключите отображение деталей. Для этого в правом нижнем углу окна Power Pivot кликните *Отображение*, и снимите галки (рис. 18.6).

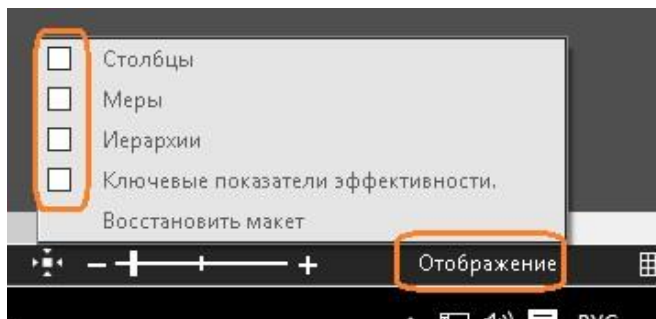


Рис. 18.6. Детали модели данных скрыты

Теперь мы можем построить единую сводную таблицу, используя меры из таблиц [Sales] и [Budget], а в качестве контекста фильтров – только поля из общих таблиц поиска. «Общими» будут те таблицы поиска, которые фильтруют обе наши таблицы данных. В нашем примере существует три общие таблицы поиска, отмеченные звездочками на рис. 18.7.

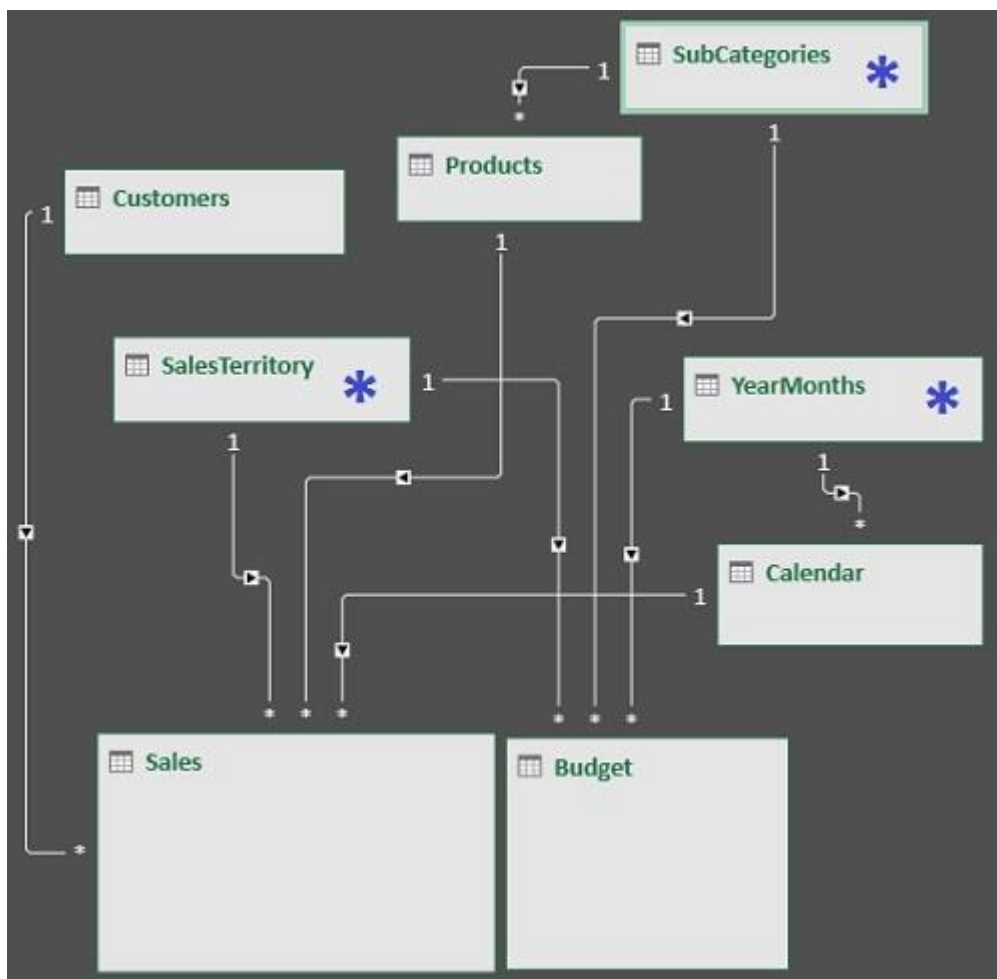


Рис. 18.7. Модель данных и направление общих контекстных фильтров

В таблице [Budget] мы создали меру [Total Budgeted Sales] = SUM(Budget[Budgeted Sales]). Поместим ее в сводную таблицу вместе с [Total Sales] из таблицы Sales:

	A	B	C	D
1				
2		Названия строк	Total Budgeted Sales	Total Sales
3		2001	\$3 236 761	\$3 266 374
4		2002	\$6 635 482	\$6 530 344
5		2003	\$10 001 112	\$9 791 060
6		2004	\$10 117 781	\$9 770 900
7		Общий итог	\$29 991 136	\$29 358 677

Рис. 18.8. Меры из разных таблиц данных: [Budget] и [Sales]

Обратите внимание, что в область *Строки* сводной вы поместите «правильное» поле Year из таблицы [YearMonths]. Если вы пометите сюда поле Year из таблицы [Calendar] у вас отобразятся неверные данные в столбце Total Budgeted Sales (рис. 18.9). Это связано с тем, что таблица [Calendar] не создает контекста фильтров для таблицы [Budget]. Чтобы уменьшить вероятность таких ошибок, рекомендую переименовывать поля, используя префиксы или суффиксы. Названия подскажут вам из каких таблиц поступают те или иные поля. В нашем примере поле *Год* из таблицы [Calendar] было названо CalendarYear.

Назван	Total Budgeted Sales	Total Sales
2001	\$29 991 136	\$3 266 374
2002	\$29 991 136	\$6 530 344
2003	\$29 991 136	\$9 791 060
2004	\$29 991 136	\$9 770 900
Общий ит	\$29 991 136	\$29 358 677

Рис. 18.9. Поле *Год* взято из неверной таблицы – [Calendar]

Гибридные меры с данными из таблиц с различной детализацией

Мы можем создать новые меры, которые ссылаются (и сравнивают) меры из таблиц Budget и Sales, несмотря на их различную детализацию.

$$[\text{Sales vs. Budget}] = ([\text{Total Sales}] - [\text{Total Budgeted Sales}]) / [\text{Total Budgeted Sales}]$$

Названия строк	Total Budgeted Sales	Total Sales	Sales vs Budget
2001	\$3 236 761	\$3 266 374	0,9%
2002	\$6 635 482	\$6 530 344	-1,5%
2003	\$10 001 112	\$9 791 060	-2,1%
2004	\$10 117 781	\$9 770 900	-3,4%
Общий итог	\$29 991 136	\$29 358 677	-2,1%

Рис. 18.10. Продажи в сравнении с бюджетом

Мы добавили условное форматирование (которое не было автоматическим). Теперь можно удалить две исходные меры, оставив только [Sales vs. Budget] и поиграть со срезами. Например, так:

Sales vs Budget	Названия строк	2001	2002	2003
Mountain Bikes				
	Australia	3,8%	-3,3%	-3,5%
	Canada	-10,0%	3,3%	-8,5%
	France	0,5%	-4,6%	-4,7%
	Germany	-15,7%	8,5%	-3,2%
	United Kingdom	7,6%	6,3%	-11,0%
	United States	3,4%	-1,7%	3,2%
Road Bikes				
	Australia	-6,5%	-1,5%	-0,1%
	Canada	-3,1%	0,7%	-6,2%
	France	-3,9%	-7,6%	0,2%
	Germany	-10,8%	-3,7%	2,4%
	United Kingdom	1,1%	4,0%	-4,9%
	United States	12,7%	-0,0%	3,9%

Рис. 18.11. Sales vs. Budget по линейкам продуктов

Использование третьего загадочного аргумента функции RANKX()

Допустим, мы получили данные о продажах конкурента: насколько хорошо их велосипеды продавались в течение последних нескольких лет (рис. 18.12). И да, информация скупа.

	ModelName	Year	SalesAmt
1	Trail 61	2001	\$4 428 670
2	Trail 412	2001	\$608 628
3	Road 187	2001	\$1 596 374
4	Mountain 385	2001	\$4 938 898
5	Trail 333	2001	\$1 655 169
6	Mountain 175	2001	\$3 290 579
7	Mountain 348	2001	\$4 487 781

Years | SubCategories | CompetitorSales | CountTest | YearM

Запись: 16 из 76

Рис. 18.12. Продажи конкурента

Задача – совместно ранжировать продажи конкурентов и собственные. Например, если одна из наших моделей показала объем продаж 3 млн. долл., а три лучшие модели конкурента – 4, 3,5 и 2,5 млн. долл., это означает, что наша модель занимает 3-е место.

Несоответствие детализации года в таблицах CompetitorSales и Sales требует создать новую таблицу подстановки – Years (рис. 18.13; напомним, что таблица данных Sales фильтруется таблицей подстановки YearMonths, в которой поле Год не является уникальным).

	Year	Добавление ст
1	2001	
2	2002	
3	2003	
4	2004	

Рис. 18.13. Новая таблица подстановки – Years

Добавим в таблицу CompetitorSales меру [Compete Sales] = SUM(CompetitorSales[SalesAmt]). А затем создадим кросс-ранговую меру в таблице Sales (или в таблице Products, если вам так больше нравится):

[Model Sales Rank vs Competition] =

```
RANKX(
    VALUES(CompetitorSales[ModelName]);
    [Compete Sales];
    [Total Sales]
)
```

Формула начинается так, как будто она собирается просто ранжировать продукты конкурентов по объему продаж (см. подробнее о [RANKX](#)). Аргумент VALUES(CompetitorSales[ModelName]) возвращает таблицу, в которой каждой модели соответствует объем продаж из таблицы CompetitorSales. Аргумент [Compete Sales] означает, что измерение, с помощью которого модели будут ранжироваться – мера [Compete Sales]. До сих пор мы имели дело с обычным использованием функции RANKX().

А вот аргумент [Total Sales] означает, что мы собираемся взять значение [Total Sales] в нашем текущем контексте фильтра (который расположен в левой колонке сводной таблицы и представлен ModelName нашей компании) и вставить его в порядок, установленный первыми двумя аргументами (рис. 18.14).

My Products' Sales			Their Sales	
Model	Total Sales	Model Sales Rank vs Competition	Model	Compete Sales
Mountain-200	\$7 929 475	14	Trail 412	\$11 064 747
Road-150	\$5 549 897	19	Trail 153	\$10 935 275
Road-250	\$4 451 260	19	Mountain 385	\$10 773 383
Touring-1000	\$2 992 008	20	Trail 487	\$10 328 574
Road-350-W	\$1 580 220	20	Trail 222	\$9 966 929
Road-550-W	\$1 514 622	20	Trail 333	\$9 955 816
Mountain-100	\$1 341 121	20	Trail 61	\$9 925 752
Road-750	\$779 206	20	Mountain 175	\$9 632 946
Road-650	\$645 380	20	Trail 493	\$8 872 714
Touring-2000	\$451 924	20	Mountain 403	\$8 648 188
Mountain-400-W	\$417 833	20	Mountain 496	\$8 262 950
Touring-3000	\$400 869	20	Road 101	\$8 220 904
Mountain-500	\$264 330	20	Road 187	\$7 955 912
Sport-100	\$225 336	20	Mountain 127	\$7 296 954
Long-Sleeve Logo Jersey	\$86 783	20	Mountain 348	\$6 882 475
Short-Sleeve Classic Jersey	\$86 168	20	Mountain 225	\$6 549 169
Women's Mountain Shorts	\$71 220	20	Trail 120	\$5 022 266

Рис. 18.14. Кросс-ранговое сравнение объемов продаж наших и конкурентных моделей. Наша топовая модель занимает 14-е место, уступая первым 13 моделям конкурента.

И, так как обе сводные таблицы фильтруются по таблице подстановки Year, мы можем добавить Год в качестве среза для обеих сводных (рис. 18.15). В 2003 году наши показатели получше!

My Products' Sales			Their Sales	
Model	Total Sales	Model Sales Rank vs Competition	Model	Compete Sales
Mountain-200	\$3 723 271	4	Trail 487	\$4 506 831
Road-250	\$2 144 214	9	Trail 333	\$3 989 343
Touring-1000	\$1 058 527	13	Road 107	\$3 971 737
Road-550-W	\$676 096	16	Trail 153	\$3 710 833
Road-350-W	\$540 915	17	Trail 412	\$3 172 682
Road-750	\$326 154	19	Mountain 403	\$3 050 286
Road-650	\$264 651	19	Trail 493	\$2 982 821
Touring-2000	\$195 591	20	Trail 329	\$2 736 560
Touring-3000	\$163 317	20	Mountain 496	\$1 722 153
Mountain-400-W	\$159 284	20	Trail 222	\$1 342 252

Рис. 18.15. Кросс-ранговое сравнение для 2003 г.